## Suraj Rimal

# **Lab Assignment: Chapter 8**

8a.

***library(ISLR)***

***set.seed(1)***

***train = sample(1:nrow(Carseats), nrow(Carseats) / 2)***

***Car.train = Carseats[train, ]***

***Car.test = Carseats[-train,]***

8b.

***library (tree)***

***reg.tree = tree(Sales~.,data = Carseats, subset=train)***

***summary(reg.tree)***

***Text

Description automatically generated***

***plot(reg.tree)***

***text(reg.tree ,pretty =0)***

***Chart

Description automatically generated***

***yh = predict(reg.tree,newdata = Car.test)***

***mean((yh - Car.test$Sales)^2)***

******

The test MSE is about 4.92.

8c.

***set.seed(1)***

***cv.car = cv.tree(reg.tree)***

***plot(cv.car$size, cv.car$dev, type = "b")***

***Chart, line chart, histogram

Description automatically generated***

We now prune the tree to obtain the 10-node tree.

***prune.car = prune.tree(reg.tree, best = 10)***

***plot(prune.car)***

***text(prune.car,pretty=0)***

Chart

Description automatically generated with low confidence

***yh=predict(prune.car, newdata= Car.test)***

***mean((yh-Car.test$Sales)^2)***

******

We see that pruning the tree slightly decreases the Test MSE to 4.918.

8d.

***library(randomForest)***

***set.seed(1)***

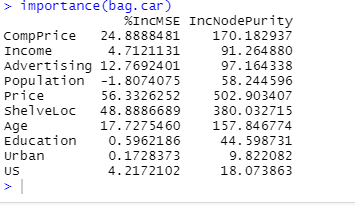
***bag.car = randomForest(Sales~.,data=Car.train,mtry = 10, importance = TRUE)***

***yh.bag = predict(bag.car,newdata=Car.test)***

***mean((yh.bag-Car.test$Sales)^2)***



***importance(bag.car)***



***varImpPlot(bag.car)***

***Chart, line chart, scatter chart

Description automatically generated***

The price that the company charges for car seats at each location, as well as the quality of the shelving location for the car seats at each location, are the most significant variables. The bagging approach regression tree's test MSE is 2.60, which is very less than that of a single tree that has been pruned optimally.

8e.

***set.seed(1)***

***rf.car = randomForest(Sales~.,data=Car.train,mtry = 3, importance = TRUE)***

***yh.rf = predict(rf.car,newdata=Car.test)***

***mean((yh.rf-Car.test$Sales)^2)***



The MSE of the test set is 2.96, when m = 3.

***set.seed(1)***

***rf.car = randomForest(Sales~.,data=Car.train,mtry = 5, importance = TRUE)***

***yh.rf = predict(rf.car,newdata=Car.test)***

***mean((yh.rf-Car.test$Sales)^2)***



slightly less when m = 5

***set.seed(1)***

***rf.car = randomForest(Sales~.,data=Car.train,mtry = 7, importance = TRUE)***

***yh.rf = predict(rf.car,newdata=Car.test)***

***mean((yh.rf-Car.test$Sales)^2)***



***set.seed(1)***

***rf.car = randomForest(Sales~.,data=Car.train,mtry = 9, importance = TRUE)***

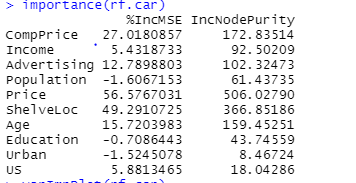
***yh.rf = predict(rf.car,newdata=Car.test)***

***mean((yh.rf-Car.test$Sales)^2)***



In this particular case, as m grows closer to 10 (the total number of predictor variables), the MSE decreases.

***importance(rf.car)***



***varImpPlot(rf.car)***

***Chart, line chart, scatter chart, box and whisker chart

Description automatically generated***

Here again, the price that the company charges for car seats at each location, as well as the quality of the shelving location for the car seats at each location, are the most significant variables.